

УДК 519.688

О прогностической способности расширенной модели Кокса: сравнительный анализ

- Микулик Илья Игоревич** — аспирант кафедры «Высшая математика». Научные интересы: прикладная математика, программирование, анализ выживаемости, методы оптимизации, параллельные вычисления. E-mail: mikulik.ilia@gmail.com
- Благовещенская Екатерина Анатольевна** — д-р физ.-мат. наук, доцент, заведующий кафедрой «Высшая математика». Научные интересы: фундаментальная и прикладная алгебра, распараллеливание алгоритмов. E-mail: kblag2002@yahoo.com

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Микулик И. И., Благовещенская Е. А. О прогностической способности расширенной модели Кокса: сравнительный анализ // Интеллектуальные технологии на транспорте. 2025. № 2 (42). С. 58–70. DOI: 10.20295/2413-2527-2025-242-58-70

Аннотация. Представлено исследование, посвященное сравнительному анализу расширенной модели Кокса с современными методами анализа выживаемости. **Цель:** проведение сравнительного анализа прогностических способностей расширенной модели Кокса с актуальными моделями и методами анализа выживаемости. Для достижения цели использованы методы машинного обучения (случайный лес выживаемости, градиентный бустинг, метод опорных векторов) и классические статистические подходы (модели Вейбулла, логлогистическая и логнормальная модели). **Методы:** анализ трех наборов данных: пациентов с раком предстательной железы, данных о рецидивах преступлений и пациентов с раком молочной железы. **Результаты:** демонстрируют, что расширенная модель Кокса превосходит или сопоставима по точности с современными методами машинного обучения, сохраняя при этом высокую интерпретируемость. **Практическая значимость:** заключается в возможности применения расширенной модели Кокса в медицине, социальных науках и других областях, где важны как точность прогнозирования, так и понимание влияния факторов на риск наступления события. Научная новизна работы заключается в проведении первого сравнительного анализа расширенной модели Кокса с другими методами анализа выживаемости, что открывает новые возможности для улучшения и адаптации модели в будущих исследованиях. Исследование имеет важное значение для развития методов анализа выживаемости и их применения в прикладных задачах, способствуя повышению точности прогнозов и улучшению интерпретируемости результатов.

Ключевые слова: анализ выживаемости, модель Кокса, метаэвристические алгоритмы, муравьиный алгоритм, оптимизация

Благодарности. Авторы благодарят к. м. н. А. Ю. Кнеева за консультацию в области медицины и онкологии в ходе работы над исследованием

1.2.2 — математическое моделирование, численные методы и комплексы программ (технические науки)

Введение

Анализ выживаемости — класс статистических моделей и методов, использующихся для оценивания вероятности наступления некоторого терминального события, после которого объект выпадает из наблюдения. Одним из ключевых преимуществ применения моделей анализа выживаемости по сравнению с традиционными регрессионными моделями является возможность работы с цензурированными данными. Цензурированные данные — это информация об объекте, собранная до определенного момента времени, после которого объект выходит из поля наблюдения. После прекращения наблюдения исследователю неизвестно, случилось ли терминальное событие для данного объекта или нет [1]. В связи с этим методы находят широкое применение в прикладных областях науки, в том числе в медицине [2].

Одним из популярных методов анализа выживаемости является полупараметрическая регрессионная модель Кокса. В работе [3] представлен метод, позволяющий строить расширенную модель Кокса — модель из класса расширения модели Кокса, не уступающую или превосходящую по прогностическим способностям классическую модель Кокса. Однако требуется провести сравнительный анализ построенных моделей с другими моделями анализа выживаемости.

Цель работы — провести сравнительный анализ прогностических способностей расширенной модели Кокса с актуальными моделями и методами анализа выживаемости. Объектом настоящего исследования является расширенная модель Кокса — обобщение модели Кокса, функция риска которой включает аддитивную и мультипликативную комбинацию ковариатов. Предметом исследования является метаэвристический алгоритм построения расширенной модели Кокса. Сравнительный анализ проводится впервые, что и является научной новизной данного исследования.

Модели и методы анализа выживаемости

Модели и методы анализа выживаемости можно разделить на классические статистические модели

и модели машинного обучения [4]. Статистические модели можно разделить на непараметрические, параметрические и полупараметрические. К непараметрическим моделям относят метод кривых Каплана — Мейера и модель Нельсона — Аалена. К параметрическим моделям относят модель Вейбулла, частным случаем которой являются экспоненциальная модель, логлогистическая модель, логнормальная модель и некоторые другие [5]. К полупараметрическим моделям относят модель пропорциональных рисков Кокса.

Непараметрические модели используются для построения кривых выживаемости. Кривые оценивают вероятность наступления терминального события по временным интервалам и фокусируются на сравнении выживаемости между группами, а не на предсказании конкретных временных исходов для объекта наблюдения. По этой причине непараметрические модели не рассматриваются в данном сравнительном анализе.

Модель Вейбулла. Функция риска модели определяется как

$$\lambda(t, x, \beta) = \alpha t^{\alpha-1} \exp(\beta x),$$

где $\alpha > 0$ — заданный коэффициент;

x — вектор значений признаков;

β — вектор влияния признаков.

При $\alpha > 1$ функция риска монотонно возрастает. В случае, когда $\alpha = 1$, модель является экспоненциальной, поэтому экспоненциальная модель является частным случаем модели Вейбулла.

Несмотря на простоту, функция риска модели Вейбулла представляет собой не линейную зависимость, а экспоненциальную, поэтому она находит применение для описания нелинейных процессов в анализе выживаемости [6].

Логлогистическая модель. Функция риска модели определяется как

$$\lambda(t, x, \beta) = \frac{\exp(\beta x) k t^{k-1}}{1 + \exp(\beta x) t^k}.$$

В логлогистической модели предполагается, что время до наступления события имеет логлогисти-

ческое распределение. Логлогистическую модель используют, когда функция риска на выборке не является монотонной. Например, в исследованиях выживаемости пациентов после определенного лечения [7].

Логнормальная модель. Она основана на предположении, что логарифм времени до события подчиняется нормальному распределению. Функция риска модели определяется как

$$\lambda(t) = \frac{f(t)}{S(t)};$$

$$f(t) = \frac{1}{\tau\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(t) - \mu)^2}{2\sigma^2}\right);$$

$$S(t) = 1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right).$$

Логнормальная модель достаточно гибкая и подходит для сложноструктурированных данных [8].

Для решения задач анализа выживаемости используют также и модели искусственного интеллекта. Методы машинного обучения нацелены на оценивание времени наступления события и дают меньше представления о характере распределения риска, а также о значимости признаков при построении прогноза.

Случайный лес выживаемости. Одним из современных и эффективных методов искусственного интеллекта, применяемых в анализе выживаемости, является случайный лес выживаемости. Этот метод представляет собой адаптацию классического алгоритма случайного леса для задач, связанных с анализом времени до наступления событий, таких как отказы оборудования, рецидивы заболеваний и другие. Основная идея метода заключается в рекурсивном разделении данных на подгруппы на основе определенного критерия разбиения, такого как максимизация различий в функциях выживаемости между группами. После разделения схожие объекты объединяются в узлы, что позволяет модели эффективно учитывать сложные зависимости между ковариатами.

Случайный лес выживаемости обладает рядом преимуществ, включая устойчивость к переобуче-

нию, способность работать с данными, содержащими пропущенные значения, и возможность обработки как категориальных, так и непрерывных переменных. Благодаря своей гибкости и высокой точности случайный лес выживаемости стал одним из наиболее популярных и широко используемых методов машинного обучения в задачах анализа выживаемости [9].

Градиентный бустинг. Является одним из наиболее популярных и гибких методов машинного обучения, который успешно применяется в задачах анализа выживаемости. Этот метод основан на идее последовательного построения ансамбля слабых моделей (обычно деревьев решений), где каждая последующая модель корректирует ошибки предыдущих. В контексте анализа выживаемости градиентный бустинг адаптируется для работы с цензурированными данными и временными событиями, что делает его особенно полезным для прогнозирования времени до наступления определенного исхода.

Одним из ключевых преимуществ градиентного бустинга является его способность учитывать сложные нелинейные зависимости между признаками и временем выживаемости. Однако у градиентного бустинга есть и свои недостатки. Во-первых, он требует значительных вычислительных ресурсов, особенно при работе с большими объемами данных. Во-вторых, интерпретируемость модели может быть затруднена из-за сложности ансамбля деревьев. Тем не менее градиентный бустинг остается одним из наиболее популярных методов в задачах анализа выживаемости, особенно в таких областях, как медицина, финансы и инженерия, где важны точные прогнозы и учет сложных взаимосвязей между переменными.

Метод опорных векторов. Для решения задач анализа выживаемости также применяется метод опорных векторов (SVM). Хотя изначально этот метод был разработан для задач классификации, он может быть модифицирован и адаптирован для решения задач регрессии, а также для анализа выживаемости [10]. В контексте анализа выживаемости метод опорных векторов использует подход, основанный на максимизации различий между класса-

ми, что позволяет модели эффективно разделять данные на группы с различными характеристиками выживаемости.

Однако, несмотря на свою универсальность, метод опорных векторов имеет ряд ограничений применительно к задачам анализа выживаемости. Основным недостатком является то, что стандартные реализации SVM не учитывают информацию об упорядоченности объектов, особенно в случаях цензурированных данных (например, когда время наступления события неизвестно для части наблюдений). Это приводит к потере важной информации о временных зависимостях и может снижать точность модели [11].

Несмотря на это, метод опорных векторов продолжает использоваться в задачах анализа выживаемости, особенно в случаях, когда данные имеют сложную структуру или требуется высокая точность классификации. Его применение особенно актуально в медицине, биоинформатике и других областях, где важна интерпретация сложных взаимосвязей между переменными [12].

Модели машинного обучения, как правило, показывают лучший результат, чем классические статистические модели, ввиду усложненной зависимости функции риска от признаков [13]. Функциональная зависимость, как правило, является «черным ящиком»: механизм работы модели скрыт от экспериментатора [14].

Модель Кокса. Одной из наиболее популярных моделей анализа выживаемости [15] является регрессионная модель пропорциональных рисков Кокса. Она относится к классу полупараметрических моделей. Функция риска для модели Кокса описывается в следующем виде:

$$\lambda(t, x, \beta) = \lambda_0(t) \exp(\beta x). \quad (1)$$

Полупараметрическая модель Кокса, включающая использование параметрического и непараметрического подходов, позволяет получить более согласованные оценки в широком диапазоне условий по сравнению с параметрическими моделями и более точные оценки по сравнению с непараметрическими методами [16]. В отличие от параметрических методов, для модели Кокса не требуется информация

о распределении времени наступления терминального события, однако предполагается, что линейная комбинация признаков коррелирует с функцией риска экспоненциально.

Важным механизмом работы регрессионной модели пропорциональных рисков Кокса является допущение о базовом риске, представленном в виде функции $\lambda_0(t)$ в функции риска. Он является функцией, зависящей от времени, и представляет собой кривую, связывающую частоту наступления терминального события и время. Важным является то, что форма функции базового риска опасности не имеет значения: она может быть возрастающей, убывающей, иметь или не иметь экстремумы. Фактически для расчета коэффициента пропорциональности Кокса не требуется знаний о форме или свойствах базового риска — она может иметь любое распределение.

Модель Кокса получила широкое распространение в медицине [17], бизнесе [18], экономике [19], биотехнологии [20], биологии [21], социальных науках [22], в исследованиях аппаратной отказоустойчивости [23].

Каждая из моделей имеет свои преимущества и недостатки. Параметрические модели не учитывают сложные зависимости между признаками и функцией риска, а сложные модели искусственного интеллекта слабо поддаются интерпретации. Также вклад признаков и функция риска могут коррелировать различным образом в зависимости от характеристик используемых данных. Это обуславливает необходимость применения методов, способных адаптироваться к специфике конкретного набора данных и строить модели анализа выживаемости, учитывающие уникальные зависимости и закономерности, присущие исследуемой выборке.

Оценка прогностической способности

Одним из наиболее распространенных показателей для оценки качества прогнозов моделей анализа выживаемости является индекс соответствия Харрелла (*c*-индекс) [24]. Этот показатель учитывает как наблюдаемые события, так и цензурированные данные, что делает его универсальным

инструментом для оценки моделей [25]. *c*-индекс количественно измеряет ранговую корреляцию между фактическим временем выживания и предсказанными значениями модели. Он рассчитывается как отношение числа правильно упорядоченных (согласованных) пар объектов к общему числу сопоставимых пар [26].

Для формального определения *c*-индекса рассмотрим случайные величины T_1 и T_2 , которые представляют время наступления терминальных событий для двух объектов из набора данных, а также \hat{T}_1 и \hat{T}_2 — соответствующие предсказанные значения времени, полученные с помощью модели. Тогда *c*-индекс может быть выражен как вероятность того, что предсказанные значения сохраняют порядок относительно реальных временных данных [27]:

$$c = P(\hat{T}_1 < \hat{T}_2 | T_1 < T_2).$$

Индекс соответствия Харрелла обладает рядом важных преимуществ:

1. Интуитивная интерпретация: значение *c*-индекса лежит в диапазоне от 0 до 1, где 1 указывает на идеальное соответствие прогнозов модели реальным данным, а значения, близкие к 0, свидетельствуют о полном несоответствии.

2. Удобство анализа: вероятностная природа индекса упрощает интерпретацию результатов. Чем выше значение *c*-индекса, тем лучше прогностические свойства модели.

3. Сравнимость: индекс позволяет сравнивать эффективность различных моделей между собой, а также оценивать пригодность наборов данных для построения моделей анализа выживаемости.

Благодаря своей универсальности и простоте интерпретации *c*-индекс широко используется в научных исследованиях, что подтверждается его применением в работах [28–30] и других.

Расширенная модель Кокса и метод ее построения

В работе [3] представлен класс моделей, функция риска которой имеет вид

$$\lambda(t, x, \beta) = \lambda_0(t) \exp(g(\beta, x)). \quad (2)$$

При $g(\beta, x) = \beta x$ функция риска становится идентичной функции риска (1) регрессионной модели Кокса. Таким образом, модель Кокса является частным случаем класса моделей, функция риска которых описывается формулой (2), поэтому представленный класс моделей называется расширением модели Кокса. Модель из класса расширения моделей называется расширенной.

Также в работе [3] предлагается использовать полиномиальную зависимость ковариатов, каждый из которых может иметь как аддитивную, так и мультипликативную связь.

Пусть $P(\hat{F})$ — полином, составленный из признаков $f \in \hat{F}$:

$$P(\hat{F}) = \sum_{i=1}^{|\xi|} \varphi_i \prod_{j=1}^{|\hat{F}|} f_j^{\xi_{ij}}, \quad (3)$$

где $\xi = \left\{ \left(\xi_1, \xi_2, \dots, \xi_{|\hat{F}|} \right)_i \right\}$, $\xi_{ij} \in \{0; 1\}$ — множество всех битовых векторов, каждый элемент которого — индикатор вхождения *j*-го признака в *i*-е слагаемое полинома,

$\varphi_i \in \{0; 1\}$ — маркер, указывающий вхождение *i*-го монома в P .

Составленный таким образом полином $P(\hat{F})$ является суммой одночленов, каждый из которых является произведением признаков. Исходя из определения (3) видно, что каждый полином однозначно описывается двумя множествами, достаточными для его построения, — ξ и \hat{F} .

$g(\hat{\beta}, X_i)$ с полиномом (3) представима в виде

$$g(\hat{\beta}, X_i) = \sum_{j=1}^{|\xi|} \hat{\beta}_j \varphi_j \prod_{k=1}^{|\hat{F}|} X_{ik}^{\xi_{jk}} | l : \hat{F}_k = F_l. \quad (4)$$

Функция риска $\lambda(t, x, \beta)$, представленная через (4), названа аддитивно-мультипликативной, так как значение каждого признака в ней может входить либо в состав суммы, либо в состав произведения. Построение функции является задачей оптимизации, которая может быть решена с помощью метаэвристических алгоритмов.

Метаэвристические алгоритмы — это обобщенные стратегии поиска решений, которые используются для решения оптимизационных задач, позволяющие находить близкие к оптимальным решения

за приемлемое в рамках задачи время. Для построения расширенной модели Кокса используется алгоритм муравьев-опылителей. Метод основан на гибридном методе муравьиной колонии и генетическом алгоритме, адаптированном под поставленную задачу. Особенностью алгоритма является преобразование набора вершин графа, представляющих признаки или их произведение в модели. Алгоритм имитирует процесс опыления и размножения цветковых растений с помощью насекомых-опылителей. Алгоритм состоит из трех частей: муравьиный алгоритм, используемый для построения модели; генетический алгоритм, используемый для улучшения работы муравьиного алгоритма; алгоритм опыления, используемый для отбора признаков или их произведения. Первые два этапа — гибридный алгоритм, представленный ниже.

Результатом работы алгоритма является расширенная модель Кокса, функция риска которой основывается на построенном полиноме $P(\hat{F})$. Каждый моном, входящий в сумму полинома, представлен в алгоритме цветком. Множество цветов образует граф, путь по которому строят муравьи-опылители. Каждый муравей составляет множество цветов, сумма соответствующих мономов которых образует полином $P(\hat{F})$. Оценкой построенного муравьем пути является фитнес-функция, которая зависит от c -индекса, а также от количества входящих в нее признаков:

$$f(S, P_q(\hat{F})) = \omega_1 c(S, P_q(\hat{F})) - \omega_2 \frac{|\hat{F}|}{p} - \omega_3 K_q - \omega_4 B_q,$$

где $c(S, P_q(\hat{F}))$ — c -индекс расширенной модели Кокса с полиномом $P_q(\hat{F})$, обученной на наборе данных S ;

p — количество признаков;

K_q, B_q — числовые характеристики, оценивающие сложность полинома;

$\omega_{1...4}$ — балансировочные коэффициенты.

Муравьиный этап алгоритма представляет собой адаптированный к задаче простой муравьиный алгоритм. Правило выбора вершины адаптировано для задачи:

$$p_v^k(t) = \frac{\tau_v^{\alpha_k}(t) \eta_v^{\beta_k}}{\sum_u \tau_u^{\alpha_k}(t) \eta_u^{\beta_k}}, \tag{5}$$

где $p_v^k(t)$ — вероятность выбора цветка v муравьем k на итерации t ;

$\tau_v(t)$ — количество отложенного феромона на цветке v на итерации t ;

η_v — эвристическая информация, которая вычисляется как $\eta_v = c(S, P_i \equiv v)$ — c -индекс расширенной модели Кокса, обученной на одном мономе цветка v .

Правило откладывания феромона также адаптировано:

$$\Delta\tau_{kv} = \frac{Q_k}{f(S, P_q(\hat{F}))},$$

где $P_q(\hat{F})$ — полином, построенный муравьем k , а f — целевая функция.

При этом правило глобального изменения феромона вместе с испарением выглядит как

$$\tau_v(t) = \rho \Delta\tau_v(t) + \sum_k \Delta\tau_{kv}. \tag{6}$$

Вторым этапом гибридного алгоритма является приложение генетического алгоритма. В данном алгоритме этот этап не отличается от генетического этапа гибридного алгоритма, кроме адаптации правила оператора селекции:

$$p_i = \frac{f(S, P_i(\hat{F}))}{\sum_j f(S, P_j(\hat{F}))}. \tag{7}$$

Феромоны, оставленные на вершинах-цветках, также используются на этапе опыления. Этап опыления реализован в виде приложения четырех операторов на популяцию цветов — селекции, кроссбридинга, лайнбридинга и старения. Каждый цветок кроме хранимого значения вершины-монома имеет параметр «возраст». Оператор селекции выбирает цветы с наибольшей концентрацией феромонов. Оператор кроссбридинга с некоторой вероятностью вводит новые цветы, моном которых представляет произведение объединения признаков из мономов цветов-родителей:

$$v_i = (e_i, \tau_i, \eta_i, o_i);$$

$$v_i v_j = v_k;$$

$$v_k = \left(\begin{array}{l} e_k = \prod_{f_q \in e_i \cup e_j} f_q, \tau_k = \frac{\tau_i + \tau_j}{2}, \\ \eta_k = c(S, P \equiv e_k), o_k = o_{\max} \end{array} \right), \quad (8)$$

где e — моном нового цветка;

τ — количество феромона, отложенного на цветок;

η — эвристическая составляющая;

o — возраст цветка;

o_{\max} — установленная продолжительность жизни цветка.

Если в результате преобразования появились цветы, уже находящиеся в популяции, то новые цветы не создаются, а обновляется возраст у имеющих цветов. Оператор лайнбридинга с небольшой вероятностью добавляет в популяцию новый цветок с единственным признаком. Этот оператор используется для того, чтобы оставить возможность вытесненным признакам участвовать в работе алгоритма. Оператор старения понижает индикатор возраста у каждого цветка. Если индикатор старения стал равен нулю, цветок выбывает из популяции.

Алгоритм можно представить в виде шагов:

Начало

1. Определить параметры $n, \tau_0, \rho, o_{\max}, \alpha_0, \beta_0, Q_0$.
2. Положить $c = 0, P = \emptyset$.
3. Положить множество цветов

$$V = \{v_i = \left(\begin{array}{l} e_i = f_i, \tau_i = rand(0, \tau_0), \\ \eta_i = c(S, P_i \equiv f_i), o_i = o_{\max} \end{array} \right) | \forall f_i \in F\}.$$

4. Положить множество муравьев

$$A = \{a_k = (\alpha_k = \alpha_0, \beta_k = \beta_0, Q_k = Q_0)\}.$$

5. Пока не достигнут критерий остановки:

5.1. Для каждого муравья $a_k \in A$:

5.1.1. $E_k(t) = \{v_{\text{random}}\}$.

5.1.2. $c_k(t-1) = 0$.

5.1.3. $c_k(t) = \eta_i$.

5.1.4. Пока $c_k(t) > c_k(t-1)$:

5.1.4.1. Выбрать v в соответствии с правилом (5)

5.1.4.2. $E_k(t) = E_k(t) \cup \{v\}$.

5.1.4.3. $c_k(t-1) = c_k(t)$.

5.1.4.4. $P_k = \sum_{i, v_i \in E_k(t)} e_i$.

5.1.4.5. $c_k(t) = f(S, P_k)$.

5.1.5. Если $c_k(t) > c$:

5.1.5.1. $c = c_k(t)$.

5.1.5.2. $P = P_k$.

5.1.6. Для каждого $v_i \in E_k(t)$ вычислить $\Delta\tau_{v_i}(t)$ в соответствии с правилом (6).

5.2. Применить оператор выбора $A = S_{\text{selection}}(A)$ правилом (7).

5.3. Применить оператор кроссинговера $A = S_{\text{crossover}}(A)$.

5.4. Применить оператор мутации $A = S_{\text{mutation}}(A)$.

5.5. Применить оператор селекции цветов $V = S_{\text{selection}}(V)$.

5.6. Применить оператор кроссбридинга $V = S_{\text{crossbreeding}}(V)$ в соответствии с правилом (8).

5.7. Применить оператор лайнбридинга $V = S_{\text{linebreeding}}(V)$.

5.8. Применить оператор старения $V = S_{\text{aging}}(V)$.

6. Вернуть значения c, P .

Критерием остановки алгоритма может являться количество итераций или сходимость решений к одному значению. Таким образом, представленный метод оптимизации муравьев-опылителей решает задачу построения аддитивно-мультипликативной функции риска и отбора признаков для расширенной модели Кокса.

Результаты

Сравнение моделей и методов анализа выживаемости проведено на трех наборах данных: базе данных больных раком предстательной железы, наборе данных о рецидивах преступлений, базе данных больных раком молочной железы. Для сравнения были представлены два экземпляра класса расширения модели Кокса: расширенная модель с наибольшим c -индексом, со всеми ковариатами и модель с выборочным числом ковариатов, функция риска которой лучше поддается интерпретации.

Результаты сравнения моделей и методов анализа выживаемости на базе больных раком предстательной железы представлены в табл. 1. База данных больных раком предстательной железы содержит клинические данные пациентов, наблюдавшихся либо получавших лечение в ФГБУ «РНЦРХТ

им. ак. А. М. Гранова» Минздрава России в период с января 1996 по декабрь 2016 года [31]. В исследование включены обезличенные данные 5073 пациентов, у которых была доступна информация о степени распространенности опухолевого процесса.

Количество признаков в построенной модели Кокса со значимыми ковариатами составило 3 из возможных 7. Исходя из результатов в табл. 1, можно сделать вывод, что построенные расширенные модели Кокса показали лучший результат в сравнении с остальными использовавшимися моделями. При этом достаточно эффективной оказалась расширенная модель Кокса, построенная всего на трех из семи признаков.

Результаты сравнения моделей на наборе данных Rossi [32] о рецидивах преступлений представлены в табл. 2. Набор данных содержит информацию о 432 заключенных, которые были выпущены из тюрем штата Мэриленд в 1970-х годах и за которыми велось наблюдение в течение года после их освобождения.

Набор данных использован для модели по нескольким причинам. Существенным преимуществом является то, что он входит в библиотеку *lifelines*, с помощью которой были построены большинство моделей в текущем анализе. Также набор является типовым, а значит, результаты могут быть сравнены с результатами в других работах. Важным является то, что в наборе нет пропуска данных. Данные не содержат большого количества признаков, что снижает временные затраты на поиск оптимальных параметров.

Количество признаков в построенной модели Кокса со значимыми ковариатами составило 3 из возможных 9. На основании данных, представленных в табл. 2, можно заключить, что наилучшая из разработанных расширенных моделей Кокса продемонстрировала незначительное превосходство по сравнению с результатами, полученными с использованием модели случайного леса для анализа выживаемости. Расширенная модель Кокса, обученная на трех признаках, показала менее высокие результаты, однако превзошла все остальные модели, включенные в анализ.

Результаты сравнения рассматриваемых моделей и методов на базе пациентов, больных раком

Таблица 1

Сравнение моделей и методов анализа выживаемости на базе больных раком предстательной железы

Модель/метод	c-индекс
Модель Кокса	0,8515
Модель Вейбулла	0,8500
Логлогистическая модель	0,8498
Логнормальная модель	0,8487
Случайный лес выживаемости	0,8517
Градиентный бустинг	0,8520
Метод опорных векторов	0,8234
Расширенная модель Кокса (все ковариаты)	0,8623
Расширенная модель Кокса (значимые ковариаты)	0,8562

Таблица 2

Сравнение моделей и методов анализа выживаемости на наборе данных о рецидивах преступлений

Модель/метод	c-индекс
Модель Кокса	0,6445
Модель Вейбулла	0,6401
Логлогистическая модель	0,6438
Логнормальная модель	0,6452
Случайный лес выживаемости	0,6649
Градиентный бустинг	0,6154
Метод опорных векторов	0,6041
Расширенная модель Кокса (все ковариаты)	0,6689
Расширенная модель Кокса (значимые ковариаты)	0,6557

молочной железы [33], представлены в табл. 3. Этот набор данных о пациентах с раком молочной железы был получен из программы SEER (Surveillance, Epidemiology, and End Results) Национального института рака (NCI) за ноябрь 2017 года, который предоставляет информацию о популяционной статистике онкологических заболеваний. Набор данных включает женщин с инфильтрирующим протоковым и дольковым раком молочной железы, диагностированным в период с 2006 по 2010 год. В исследование было включено 4024 пациента.

Количество признаков в построенной модели Кокса со значимыми ковариатами составило 6 из возможных 8. На основании данных, представленных в табл. 3, можно заключить, что наилучшая из разработанных расширенных моделей Кокса незначительно уступила методу опорных векторов. Расширенная модель Кокса, обученная на шести признаках, также уступила только этому методу, однако превзошла все остальные модели, рассмотренные в анализе.

Таблица 3

**Сравнение моделей и методов анализа
выживаемости на базе данных больных раком
молочной железы**

Модель/метод	c-индекс
Модель Кокса	0,6910
Модель Вейбулла	0,7025
Логлогистическая модель	0,7026
Логнормальная модель	0,7028
Случайный лес выживаемости	0,6996
Градиентный бустинг	0,7237
Метод опорных векторов	0,7281
Расширенная модель Кокса (все ковариаты)	0,7240
Расширенная модель Кокса (значимые ковариаты)	0,7238

Проведенное сравнение расширенных моделей Кокса с другими методами анализа выживаемости на трех наборах данных показало их высокую эффективность. На данных о раке предстательной железы и рецидивах преступлений расширенные модели Кокса, включая версии с ограниченным числом признаков, продемонстрировали превосходство над большинством альтернативных методов. На данных о раке молочной железы модель Кокса немного уступила методу опорных векторов, но осталась одной из наиболее результативных. Таким образом, расширенные модели Кокса подтвердили свою универсальность и высокую предсказательную способность, особенно при интерпретируемости функции риска.

Заключение

В статье проведен сравнительный анализ прогностических способностей расширенной модели Кокса с другими современными методами анализа выжи-

ваемости. Расширенная модель Кокса, представляющая собой обобщение классической модели Кокса, построена с использованием метаэвристического алгоритма муравьев-опылителей. Особенностью этих моделей является включение в функцию риска как аддитивных, так и мультипликативных комбинаций ковариатов. Это, с одной стороны, обеспечивает высокую точность прогнозирования, с другой — сохраняет интерпретируемость результатов.

Результаты исследования, проведенного на трех различных наборах данных (рак предстательной железы, рецидивы преступлений и рак молочной железы), показали, что расширенная модель Кокса демонстрирует конкурентоспособные результаты в сравнении с классическими статистическими моделями анализа выживаемости и методами машинного обучения. В частности, на данных о раке предстательной железы и рецидивах преступлений расширенная модель Кокса превзошла большинство альтернативных методов. На данных о раке молочной железы расширенная модель Кокса показала результаты, близкие к методу опорных векторов, сохранив при этом свою интерпретируемость, что подтверждает ее универсальность и возможность применения в прикладных областях.

Таким образом, расширенная модель Кокса представляет собой эффективный инструмент для анализа выживаемости, сочетающий в себе высокую прогностическую способность и интерпретируемость. Результаты исследования подтверждают возможность дальнейшего развития и применения метода построения данной модели в прикладных задачах, особенно в медицине и социальных науках.

СПИСОК ИСТОЧНИКОВ

1. Летчиков А. В., Матвеев Р. Ю., Широкова М. А. Решение проблемы цензурированных данных при моделировании оценки индивидуального кредитного риска // Вестник Удмуртского университета. Серия «Экономика и право». 2019. Т. 29, Вып. 1. С. 34–41.
2. Heterogeneous Datasets for Federated Survival Analysis Simulation / A. Archetti, E. Lomurno, F. Lattari [et al.] // ICPE '23 Companion: Companion of the ACM/SPEC International Conference on Performance Engineering (Coimbra, Portugal, 15–19 April 2023). New York (NY): Association for Computing Machinery, 2023. Pp. 173–180. DOI: 10.1145/3578245.3584935.
3. Микулик И. И., Жаринов Г. М., Кнеев А. Ю. Алгоритм построения функции риска расширенной модели Кокса и его применение на базе данных больных раком предстательной железы // Advanced Engineering Research (Rostov-on-Don). 2024. Т. 24, № 4. С. 413–423. DOI: 10.23947/2687-1653-2024-24-4-413-423.
4. Wang P., Li Y., Reddy C. K. Machine Learning for Survival Analysis: A Survey // ACM Computing Surveys. 2019. Vol. 51, Iss. 6. Art. No. 110. 36 p. DOI: 10.1145/3214306.

5. Overview of Parametric Survival Analysis for Health-Economic Applications / K. J. Ishak, N. Kreif, A. Benedict, N. Muszbek // *PharmacoEconomics*. 2013. Vol. 31, Iss. 8. Pp. 663–675. DOI: 10.1007/s40273-013-0064-3.
6. Zhu S., Campanella O., Chen G. Estimation of Parameters in the Weibull Model from Microbial Survival Data Obtained Under Constant Conditions with Come-up Times // *Journal of Food Engineering*. 2021. Vol. 292. Art. No. 110364. 10 p. DOI: 10.1016/j.jfoodeng.2020.110364.
7. Borges P. Estimating the Turning Point of the Log-Logistic Hazard Function in the Presence of Long-Term Survivors with an Application for Uterine Cervical Cancer Data // *Journal of Applied Statistics*. 2021. Vol. 48, Iss. 2. Pp. 203–213. DOI: 10.1080/02664763.2020.1720627.
8. Extrapolation of Survival Curves Using Standard Parametric Models and Flexible Parametric Spline Models: Comparisons in Large Registry Cohorts with Advanced Cancer / J. Gray, T. Sullivan, N. R. Latimer [et al.] // *Medical Decision Making*. 2021. Vol. 41, Iss. 2. Pp. 179–193. DOI: 10.1177/0272989X20978958.
9. Decision Tree for Competing Risks Survival Probability in Breast Cancer Study / N. A. Ibrahim, A. Kudus, I. Daud, M. R. Abu Bakar // *International Journal of Biological and Medical Sciences*. 2008. Vol. 3, Iss. 1. Pp. 25–29.
10. Support Vector Machines for Survival Analysis / V. Van Belle, K. Pelckmans, J. A. K. Suykens, S. Van Huffel // *Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2007)*, (Plymouth, United Kingdom, 25–27 July 2007). 8 p.
11. Shivaswamy P. K., Chu W., Jansche M. A Support Vector Approach to Censored Targets // *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007)* (Omaha, NE, USA, 28–31 October 2007). Institute of Electrical and Electronics Engineers, 2007. Pp. 655–660. DOI: 10.1109/ICDM.2007.93.
12. Integrating Bioinformatics and Machine Learning Methods to Analyze Diagnostic Biomarkers for HBV-induced Hepatocellular Carcinoma / A. Yang, J. Liu, M. Li [et al.] // *Diagnostic Pathology*. 2024. Vol. 19, Iss. 1. Art. No. 105. 10 p. DOI: 10.1186/s13000-024-01528-8.
13. Deep Learning for Survival Analysis: A Review / S. Wiegrecbe, P. Kopper, R. Sonabend [et al.] // *Artificial Intelligence Review*. 2024. Vol. 57, Iss. 3. Art. No. 65. 34 p. DOI: 10.1007/s10462-023-10681-3.
14. Kovalev M. S., Utkin L. V. A Robust Algorithm for Explaining Unreliable Machine Learning Survival Models Using the Kolmogorov — Smirnov Bounds // *Neural Networks*. 2020. Vol. 132. Pp. 1–18. DOI: 10.1016/j.neunet.2020.08.007.
15. Archetti A., Ieva F., Matteucci M. Scaling Survival Analysis in Healthcare with Federated Survival Forests: A Comparative Study on Heart Failure and Breast Cancer Genomics // *Future Generation Computer Systems*. 2023. Vol. 149. Pp. 343–358. DOI: 10.1016/j.future.2023.07.036.
16. Powell J. L. Estimation of Semiparametric Models // *Handbook of Econometrics*. Volume IV / R. F. Engle, D. L. McFadden (eds.). Amsterdam: North-Holland Publishing, 1994. Pp. 2443–2521. DOI: 10.1016/S1573-4412(05)80010-8.
17. Assessing Performance and Clinical Usefulness in Prediction Models with Survival Outcomes: Practical Guidance for Cox Proportional Hazards Models / D. J. McLernon, D. Giardiello, B. Van Calster // *Annals of Internal Medicine*. 2023. Vol. 176, No. 1. Pp. 105–114. DOI: 10.7326/M22-0844.
18. Tao X., Wang M., Ji Y. The Application of Graph-Structured Cox Model in Financial Risk Early Warning of Companies // *Sustainability*. 2023. Vol. 15, Iss. 14. Art. No. 10802. 16 p. DOI: 10.3390/su151410802.
19. Gomez-Gonzalez J. E., Uribe J. M., Valencia O. M. Does Economic Complexity Reduce the Probability of a Fiscal Crisis? // *World Development*. 2023. Vol. 168. Art. No. 106250. 17 p. DOI: 10.1016/j.worlddev.2023.106250.
20. Viral Kinetics in Sylvatic Yellow Fever Cases / V. I. Avelino-Silva, M. V. Thomazella, M. P. Marmorato [et al.] // *The Journal of Infectious Diseases*. 2023. Vol. 227, Iss. 9. Pp. 1097–1103. DOI: 10.1093/infdis/jiac435.
21. Survival Analysis of Productive Life in Florida Dairy Goats Using a Cox Proportional Hazards Model / C. Ziadi, J. P. Sánchez, M. Sánchez [et al.] // *Journal of Animal Breeding and Genetics*. 2023. Vol. 140, Iss. 4. Pp. 431–439. DOI: 10.1111/jbg.12769.
22. High Social Risk and Mortality. A Prospective Study in Community-Dwelling Older Adults Living in a Rural Ecuadorian Village / O. H. Del Brutto, R. M. Mera, D. A. Rumbica [et al.] // *Preventive Medicine Reports*. 2023. Vol. 32. Art. No. 102146. 4 p. DOI: 10.1016/j.pmedr.2023.102146.
23. Papathanasiou D., Demertzis K., Tziritas N. Machine Failure Prediction Using Survival Analysis // *Future Internet*. 2023. Vol. 15, Iss. 5. Art. No. 153. 26 p. DOI: 10.3390/fi15050153.
24. Evaluating the Yield of Medical Tests / F. E. Harrell Jr., R. M. Califf, D. B. Pryor [et al.] // *JAMA: The Journal of the American Medical Association*. 1982. Vol. 247, No. 18. Pp. 2543–2546. DOI: 10.1001/jama.1982.03320430047030.
25. The Concordance Index Decomposition: A Measure for a Deeper Understanding of Survival Prediction Models / A. Alabdallah, M. Ohlsson, S. Pashami, T. Rögnvaldsson // *Artificial Intelligence in Medicine*. 2024. Vol. 148. Art. No. 102781. 10 p. DOI: 10.1016/j.artmed.2024.102781.

26. Pitfalls of the Concordance Index for Survival Outcomes / N. Hartman, S. Kim, K. He, J. D. Kalbfleisch // *Statistics in Medicine*. 2023. Vol. 42, Iss. 13. Pp. 2179–2190. DOI: 10.1002/sim.9717.
27. Румянцева Е. В., Фурманов К. К. Использование вневыборочных остатков Кокса — Снелл при прогнозировании наступления событий // *Бизнес-информатика*. 2021. Т. 15, № 1. С. 7–18. DOI: 10.17323/2587-814X.2021.1.7.18.
28. Longato E., Vettoretti M., Di Camillo B. A Practical Perspective on the Concordance Index for the Evaluation and Selection of Prognostic Time-to-Event Models // *Journal of Biomedical Informatics*. 2020. Vol. 108. Art. No. 103496. 9 p. DOI: 10.1016/j.jbi.2020.103496.
29. Pencina M. J., D’Agostino R. B. Overall C as a Measure of Discrimination in Survival Analysis: Model Specific Population Value and Confidence Interval Estimation // *Statistics in Medicine*. 2004. Vol. 23, Iss. 13. Pp. 2109–2123. DOI: 10.1002/sim.1802.
30. A Convolutional Neural Network Model for Survival Prediction Based on Prognosis-Related Cascaded Wx Feature Selection / Q. Yin, W. Chen, C. Zhang, Z. Wei // *Laboratory Investigation*. 2022. Vol. 102, Iss. 10. Pp. 1064–1074. DOI: 10.1038/s41374-022-00801-y.
31. Свидетельство о государственной регистрации базы данных № 2016620331 Российская Федерация. База данных больных раком предстательной железы: опублик. 20.04.2016 / Г. М. Жаринов.
32. Rossi P. H., Berk R. A., Lenihan K. J. *Money, Work, and Crime: Experimental Evidence*. New York: Academic Press, 1980. 260 p.
33. Teng J. SEER Breast Cancer Data // *IEEE Dataport*. Last update 18.01.2019. DOI: 10.21227/a9qy-ph35.

Дата поступления: 03.03.2025

Решение о публикации: 07.03.2025

Comparative Analysis of the Predictive Ability of the Extended Cox Model

Ilya I. Mikulik

— Postgraduate Student of the Higher Mathematics Department. Research interests: applied mathematics, programming, survival analysis, optimization methods, parallel computing. E-mail: mikulik.ilia@gmail.com

Ekaterina A. Blagoveshchenskaya

— Dr. Sci. in Physics and Mathematics, Associate Professor, the Head of the Higher Mathematics Department. Research interests: fundamental and applied algebra, parallelization of algorithms. E-mail: kblag2002@yahoo.com

Emperor Alexander I St. Petersburg State Transport University, 9, Moskovsky ave., Saint Petersburg, 190031, Russia

For citation: Mikulik I. I., Blagoveshchenskaya E. A. Comparative Analysis of the Predictive Ability of the Extended Cox Model. *Intellectual Technologies on Transport*, 2025, No. 2 (42), Pp. 58–70. DOI: 10.20295/2413-2527-2025-242-58-70. (In Russian)

Abstract. *A study presents a comparative analysis of the extended Cox model with modern survival analysis methods. **Purpose:** to evaluate the predictive abilities of the extended Cox model in comparison with the current survival analysis models and techniques. To achieve this goal, machine learning methods (survival random forest, gradient boosting, support vector machines) and classical statistical approaches (Weibull, log-logistic, and log-normal models) were used. **Methods:** analyzing three datasets, that is prostate cancer patients, data on their recurrent admissions, and breast cancer patients. **Results:** To demonstrate that the extended Cox model outperforms or is comparable in accuracy to modern machine learning methods while maintaining high interpretability. **Practical significance:** the applicability of the extended Cox model in medicine, social sciences, and other fields where both prediction accuracy and understanding of the factors affecting the risk of an event are crucial. The scientific novelty of this study lies in conducting a first comparative analysis of the extended Cox model with other survival analysis methods opening new opportunities for improving and*

adapting the model in future research. The study will be of great importance for the development of survival analysis methods and their application in practical tasks contributing to increased prediction accuracy and improved interpretability of results.

Keywords: *survival analysis, Cox model, metaheuristic algorithms, ant colony algorithm, optimization*

REFERENCES

1. Letchikov A. V., Matveev R. Yu., Shirobokova M. A. Reshenie problemy tsenzurovannykh dannykh pri modelirovanii otsenki individualnogo kreditnogo riska [Solving the Problem of Censored Data in Modeling the Individual Credit Risk Estimation], *Vestnik Udmurtskogo universiteta. Seriya "Ekonomika i pravo"* [The Bulletin of Udmurt University. Series Economics and Law], 2019, Vol. 29, Iss. 1, Pp. 34–41. (In Russian)
2. Archetti A., Lomurno E., Lattari F., et al. Heterogeneous Datasets for Federated Survival Analysis Simulation, *ICPE '23 Companion: Companion of the ACM/SPEC International Conference on Performance Engineering (Coimbra, Portugal, April 15–19, 2023)*. New York (NY), Association for Computing Machinery, 2023, Pp. 173–180. DOI: 10.1145/3578245.3584935.
3. Mikulik I. I., Zharinov G. M., Kneev A. Yu. Algorithm for Constructing the Hazard Function of the Extended Cox Model and Its Application to the Prostate Cancer Patient Database, *Advanced Engineering Research (Rostov-on-Don)*, 2024, Vol. 24, No. 4, Pp. 413–423. DOI: 10.23947/2687-1653-2024-24-4-413-423.
4. Wang P., Li Y., Reddy C. K. Machine Learning for Survival Analysis: A Survey, *ACM Computing Surveys*, 2019, Vol. 51, Iss. 6, Art. No. 110, 36 p. DOI: 10.1145/3214306.
5. Ishak K. J., Kreif N., Benedict A., Muszbek N. Overview of Parametric Survival Analysis for Health-Economic Applications, *PharmacoEconomics*, 2013, Vol. 31, Iss. 8, Pp. 663–675. DOI: 10.1007/s40273-013-0064-3.
6. Zhu S., Campanella O., Chen G. Estimation of Parameters in the Weibull Model from Microbial Survival Data Obtained Under Constant Conditions with Come-up Times, *Journal of Food Engineering*, 2021, Vol. 292, Art. No. 110364, 10 p. DOI: 10.1016/j.jfoodeng.2020.110364.
7. Borges P. Estimating the Turning Point of the Log-Logistic Hazard Function in the Presence of Long-Term Survivors with an Application for Uterine Cervical Cancer Data, *Journal of Applied Statistics*, 2021, Vol. 48, Iss. 2, Pp. 203–213. DOI: 10.1080/02664763.2020.1720627.
8. Gray J., Sullivan T., Latimer N. R., et al. Extrapolation of Survival Curves Using Standard Parametric Models and Flexible Parametric Spline Models: Comparisons in Large Registry Cohorts with Advanced Cancer, *Medical Decision Making*, 2021, Vol. 41, Iss. 2, Pp. 179–193. DOI: 10.1177/0272989X20978958.
9. Ibrahim N. A., Kudus A., Daud I., Abu Bakar M. R. Decision Tree for Competing Risks Survival Probability in Breast Cancer Study, *International Journal of Biological and Medical Sciences*, 2008, Vol. 3, Iss. 1, Pp. 25–29.
10. Van Belle V., Pelckmans K., Suykens J. A. K., Van Huffel S. Support Vector Machines for Survival Analysis, *Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2007), Plymouth, United Kingdom, July 25–27, 2007*. 8 p.
11. Shivaswamy P. K., Chu W., Jansche M. A Support Vector Approach to Censored Targets, *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007), Omaha, NE, USA, October 28–31, 2007*. Institute of Electrical and Electronics Engineers, 2007, Pp. 655–660. DOI: 10.1109/ICDM.2007.93.
12. Yang A., Liu J., Li M., et al. Integrating Bioinformatics and Machine Learning Methods to Analyze Diagnostic Biomarkers for HBV-induced Hepatocellular Carcinoma, *Diagnostic Pathology*, 2024, Vol. 19, Iss. 1, Art. No. 105, 10 p. DOI: 10.1186/s13000-024-01528-8.
13. Wiegrefe S., Kopper P., Sonabend R., et al. Deep Learning for Survival Analysis: A Review, *Artificial Intelligence Review*, 2024, Vol. 57, Iss. 3, Art. No. 65, 34 p. DOI: 10.1007/s10462-023-10681-3.
14. Kovalev M. S., Utkin L. V. A Robust Algorithm for Explaining Unreliable Machine Learning Survival Models Using the Kolmogorov — Smirnov Bounds, *Neural Networks*, 2020, Vol. 132, Pp. 1–18. DOI: 10.1016/j.neunet.2020.08.007.
15. Archetti A., Ieva F., Matteucci M. Scaling Survival Analysis in Healthcare with Federated Survival Forests: A Comparative Study on Heart Failure and Breast Cancer Genomics, *Future Generation Computer Systems*, 2023, Vol. 149, Pp. 343–358. DOI: 10.1016/j.future.2023.07.036.

16. Powell J. L. Estimation of Semiparametric Models. In: *Engle R. F., McFadden D. L. (eds.) Handbook of Econometrics. Volume IV.* Amsterdam, North-Holland Publishing, 1994, Pp. 2443–2521. DOI: 10.1016/S1573-4412(05)80010-8.
17. McLernon D. J., Giardiello D., Van Calster B. Assessing Performance and Clinical Usefulness in Prediction Models with Survival Outcomes: Practical Guidance for Cox Proportional Hazards Models, *Annals of Internal Medicine*, 2023, Vol. 176, No. 1, Pp. 105–114. DOI: 10.7326/M22-0844.
18. Tao X., Wang M., Ji Y. The Application of Graph-Structured Cox Model in Financial Risk Early Warning of Companies, *Sustainability*, 2023, Vol. 15, Iss. 14, Art. No. 10802, 16 p. DOI: 10.3390/su151410802.
19. Gomez-Gonzalez J. E., Uribe J. M., Valencia O. M. Does Economic Complexity Reduce the Probability of a Fiscal Crisis? *World Development*, 2023, Vol. 168, Art. No. 106250, 17 p. DOI: 10.1016/j.worlddev.2023.106250.
20. Avelino-Silva V. I., Thomazella M. V., Marmorato M. P., et al. Viral Kinetics in Sylvatic Yellow Fever Cases, *The Journal of Infectious Diseases*, 2023, Vol. 227, Iss. 9, Pp. 1097–1103. DOI: 10.1093/infdis/jiac435.
21. Ziadi C., Sánchez J. P., Sánchez M., et al. Survival Analysis of Productive Life in Florida Dairy Goats Using a Cox Proportional Hazards Model, *Journal of Animal Breeding and Genetics*, 2023, Vol. 140, Iss. 4, Pp. 431–439. DOI: 10.1111/jbg.12769.
22. Del Brutto O. H., Mera R. M., Rumbela D. A., et al. High Social Risk and Mortality. A Prospective Study in Community-Dwelling Older Adults Living in a Rural Ecuadorian Village, *Preventive Medicine Reports*, 2023, Vol. 32, Art. No. 102146, 4 p. DOI: 10.1016/j.pmedr.2023.102146.
23. Papathanasiou D., Demertzis K., Tziritas N. Machine Failure Prediction Using Survival Analysis, *Future Internet*, 2023, Vol. 15, Iss. 5, Art. No. 153, 26 p. DOI: 10.3390/fi15050153.
24. Harrell Jr. F. E., Califf R. M., Pryor D. B., et al. Evaluating the Yield of Medical Tests, *JAMA: The Journal of the American Medical Association*, 1982, Vol. 247, No. 18, Pp. 2543–2546. DOI: 10.1001/jama.1982.03320430047030.
25. Alabdallah A., Ohlsson M., Pashami S., Rögnvaldsson T. The Concordance Index Decomposition: A Measure for a Deeper Understanding of Survival Prediction Models, *Artificial Intelligence in Medicine*, 2024, Vol. 148, Art. No. 102781, 10 p. DOI: 10.1016/j.artmed.2024.102781.
26. Hartman N., Kim S., He K., Kalbfleisch J. D. Pitfalls of the Concordance Index for Survival Outcomes, *Statistics in Medicine*, 2023, Vol. 42, Iss. 13, Pp. 2179–2190. DOI: 10.1002/sim.9717.
27. Rumyantseva E. V., Furmanov K. K. Using Out-of-Sample Cox — Snell Residuals in Time-to-Event Forecasting, *Business Informatics*, 2021, Vol. 15, No. 1, Pp. 7–18. DOI: 10.17323/2587-814X.2021.1.7.18.
28. Longato E., Vettoretti M., Di Camillo B. A Practical Perspective on the Concordance Index for the Evaluation and Selection of Prognostic Time-to-Event Models, *Journal of Biomedical Informatics*, 2020, Vol. 108, Art. No. 103496, 9 p. DOI: 10.1016/j.jbi.2020.103496.
29. Pencina M. J., D’Agostino R. B. Overall C as a Measure of Discrimination in Survival Analysis: Model Specific Population Value and Confidence Interval Estimation, *Statistics in Medicine*, 2004, Vol. 23, Iss. 13, Pp. 2109–2123. DOI: 10.1002/sim.1802.
30. Yin Q., Chen W., Zhang C., Wei Z. A Convolutional Neural Network Model for Survival Prediction Based on Prognosis-Related Cascaded Wx Feature Selection, *Laboratory Investigation*, 2022, Vol. 102, Iss. 10, Pp. 1064–1074. DOI: 10.1038/s41374-022-00801-y.
31. Zharinov G. M. Baza dannykh bolnykh rakom predstatelnoy zhelezy [Prostate cancer patient database]. Certificate of State registration of the database RU No. 2016620331, published at April 20, 2016. (In Russian)
32. Rossi P. H., Berk R. A., Lenihan K. J. Money, Work, and Crime: Experimental Evidence. New York, Academic Press, 1980. 260 p.
33. Teng J. SEER Breast Cancer Data, *IEEE Dataport*. Last update January 18, 2019. DOI: 10.21227/a9qy-ph35.

Received: 03.03.2025

Accepted: 07.03.2025